Parametric Sensitivity Analysis for Biochemical Reaction Networks based on Pathwise Information Theory

Yannis Pantazis[1], Markos A. Katsoulakis*[1] and Dionisios G. Vlachos[2]

[1]Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA, 01002. USA
[2]Department of Chemical Engineering, University of Delaware, Newark, Delaware, 19716. USA

Email: Yannis Pantazis - pantazis@math.umass.edu; Markos A. Katsoulakis*- markos@math.umass.edu; Dionisios G. Vlachos - vlachos@udel.edu;

*Corresponding author

Abstract

Background: Stochastic modeling and simulation provide powerful predictive methods for the intrinsic understanding of fundamental mechanisms in complex biochemical networks. Typically such mathematical models involve networks of coupled jump stochastic processes with a large number of parameters that need to be suitably calibrated against experimental data. In this direction, the sensitivity analysis of the reaction network's parameters is a crucial mathematical and computational tool since it allows to infer information about the robustness and the identifiability of the model's parameters. However, existing sensitivity analysis approaches such as variants of the finite difference method can have an overwhelming computational cost in models with a high-dimensional parameter space.

Results: We develop a sensitivity analysis methodology suitable for complex stochastic reaction networks with a high number of parameters. The proposed approach is based on Information Theory and relies on the quantification of information loss due to parameter perturbations between time-series distributions, hence referred to as "pathwise". This is achieved by employing the rigorously-derived Relative Entropy Rate (RER), which is directly computable from the propensity functions. A key aspect of the method is that an associated pathwise Fisher Information Matrix (FIM) is defined, which in turn constitutes a gradient-free approach to quantify the parameter sensitivities. The study of the structure of the FIM which turns out to be block-diagonal reveals hidden parameter dependencies and sensitivities in reaction networks.

Conclusions: As a gradient-free method, the proposed sensitivity analysis provides a significant advantage when dealing with complex stochastic systems with a large number of parameters. In addition, the knowledge of the structure of the FIM can allow to efficiently address questions on parameter identifiability, estimation and robustness. The proposed method is tested and validated on two biochemical systems, namely: (a) the p53 reaction network where quasi-steady stochastic oscillations of the concentrations are observed, and for which continuum approximations (e.g. mean field, stochastic Langevin, etc.) break down due to persistent oscillations between high and low populations, and (b) an Epidermal Growth Factor Receptor (EGFR) model which is an example of a high-dimensional stochastic reaction network with more than 200 reactions and a corresponding number of parameters.

Keywords: Biochemical reaction networks, sensitivity analysis, relative entropy rate, Fisher information matrix, p53 model, EGFR model

1 Background

The need of an intrinsic understanding of the interplay between complexity and robustness of biological processes and their corresponding design principles is well-documented, see for instance [1–5]. The concept of robustness can be described as "a property that allows a system to maintain its functions against internal and external perturbations" [3]. When referring to mathematical models of complex biological processes, one of the mathematical tools to describe the robustness of a system to perturbations is sensitivity analysis which attempts to determine which parameter directions (or their combinations) are the most/least sensitive to perturbations and uncertainty or to errors resulting from experimental parameter estimation. Recently there has been significant progress in developing sensitivity analysis tools for low-dimensional stochastic processes, modeling well-mixed chemical reactions and biological networks. Some of the mathematical tools included log-likelihood methods and Girsanov transformations [6–8], polynomial chaos [9], finite difference methods and their variants [10, 11] and pathwise sensitivity methods [12]. However, existing sensitivity analysis approaches can have an overwhelming computational cost, either due to high variance in the gradient estimators, or in models with a high-dimensional parameter space. Such issues and comparisons between methods where discussed in [13].

The aforementioned methods focus on the sensitivity of stochastic trajectories and corresponding averages. However, it is often the case that we are interested in the sensitivity of probability density functions (PDF), which in nonlinear and/or discrete systems are typically non-Gaussian. In that latter direction, there is a broad recent literature relying on information theory tools, and where sensitivity is estimated by using the Relative Entropy and the Fisher Information Matrix between PDFs, providing a quantification of information loss along different parameter perturbations. We refer to [14–18] for the case when the parametric PDF is explicitly known. For instance, in [16] the parametric PDF's structure is known as it is obtained through an entropy maximization subject to constraints. Knowing the form of the PDF allows to carry out calculations such as estimating the relative entropy and identifying the most sensitive parameter combinations. Furthermore, the pathwise PDFs are also known in reaction networks when a Linear Noise Approximation (LNA) is employed and in which case the relative entropy can be explicitly computed allowing thus to carry out a sensitivity analysis, [18]. However, for complex stochastic dynamics of large reaction networks, spatial Kinetic Monte Carlo algorithms and molecular dynamics, such explicit formulas for the PDFs are in general not explicitly available.

In [19] we address such challenges, by introducing a new methodology for complex stochastic dynamics based on the Relative Entropy Rate (RER) which provides a measure of the sensitivity of the entire time-series distribution. Typically the space of all such time-series is referred in probability theory as the "path-space". RER measures the loss of information per unit time in path space after an arbitrary perturbation of parameter combinations. RER and the corresponding Fisher Information Matrix (FIM) become computationally feasible as they admit explicit formulas which depend only on the propensity functions (see (6) and (8), respectively). In fact, in [19] we showed that the proposed approach to sensitivity analysis has the following features: First, it is rigorously valid for the sensitivity of long-time, stationary dynamics in path space, including for example bistable, periodic and pulse-like dynamics. Second, it is a gradient-free sensitivity analysis method suitable for high-dimensional parameter spaces as the ones typically arising in complex biochemical networks. Third, the RER method does not require the explicit knowledge of the equilibrium PDFs, relying only on information for local dynamics making it suitable for non-equilibrium steady state systems. In [19] we demonstrated these features by focusing on two classes of problems: Langevin particle systems with either reversible (gradient) or non-reversible (non-gradient) forcing, highlighting the ability of the method to carry out sensitivity analysis in non-equilibrium systems; and spatially extended Kinetic Monte Carlo models, showing that the method can handle high-dimensional problems.

In this paper, we extend and apply the pathwise sensitivity analysis method in [19] to biochemical reaction networks, and demonstrate the intrinsic sensitivity structure of the network. Such systems are typically modeled as population stochastic jump Markov processes and they are simulated using either exact algorithms such as the Stochastic Simulation Algorithm (SSA), [20–22] and the next-reaction method [23] or by employing approximations such as mean field ODEs, tau-leap [24] and stochastic Langevin methods [25]. We show that the proposed pathwise method allows us to discover the intrinsic sensitivities of the reaction network by decomposing the FIM into diagonal blocks. The block-diagonal structure of the proposed FIM reveals, in a straightforward way, the sensitivity interdependencies between the system parameters. For instance, if each propensity function depends only on one parameter –usually the reaction constant– then the FIM is a diagonal matrix (see (15)). The sparse representation of the FIM can be essential in optimal experimental design as well as in parameter identifiability and robustness where each subset of the parameters defined by a block of the FIM can be treated separately. Moreover, our earlier rigorous analysis [19] for the stationary process regime suggests suitable extensions in the transient case which are here tested and validated. Finally, we present strategies for efficiently and reliably implementing the proposed method for high-dimensional, complex stochastic systems using an array of existing accelerated versions of the SSA algorithm such as mean field, stochastic Langevin, τ-leap approximations and their variants, [21, 24–27].

We test the proposed set of methods and computational strategies in two examples of complex biochemical networks. First, we study the parameter sensitivities of a p53 gene model for cell cycle regulation and response to DNA damage, that incorporates the feedback between the tumor suppressor p53 and the oncogene Mdm2 [28]. This is a reaction network that exhibits random oscillations in its steady state, and for which continuum approximations of the SSA such as LNA break down due to persistent oscillations between high and low populations. Using the proposed method, we also study a far more complex network, the epidermal growth factor receptor (EGFR) model, describing signaling processes between mammalian cells [29–31]. This is a high-dimensional system both in the number of variables and parameters, including 94 species and 207 reactions. Having a gradient-free method such as FIM for this example with parameter space of dimension 207 provides a significant advantage over gradient methods such as finite differencing, where the computation of a very high number of partial derivatives and/or directional derivatives is needed and with possibly significant variance that scales with the dimension, [11]. By contrast, the eigenvalue/eigenvector analysis of the proposed FIM identifies the order from least to most sensitive directions (determined by the eigenvectors of the FIM) by correspondingly ordering the eigenvalues.

In Methods we present the derivation of the Relative Entropy Rate and its corresponding Fisher Information Matrix for continuous-time jump Markov processes as well as we reveal the block-diagonal structure of the FIM for commonly encountered reaction networks, continued by the presentation of both unbiased and biased –but accelerated– statistical estimators for RER and FIM. Then in the Results we applied and validated the proposed pathwise sensitivity analysis methodology in two complex biological reaction networks.

Methods

We consider a well-mixed reaction network with N species, $\mathbf{S} = \{S_1, ..., S_N\}$, and M reactions, $\mathbf{R} = \{R_1, ..., R_M\}$. The state of the system at any time $t \geq 0$ is denoted by an N-dimensional vector $\mathbf{X}(t) = [X_1(t), ..., X_N(t)]^T$ where $X_i(t)$ is the number of molecules of species S_i at time t. Let the N-dimensional vector v_j correspond to the stoichiometry vector of j-th reaction such that $v_{i,j}$ is the stoichiometric coefficient of species S_i in reaction R_j. Given that the reaction network at time t is in state $\mathbf{X}(t) = \mathbf{x}$, the propensity function, $a_j(\mathbf{x})$, is defined such that the infinitesimal quantity $a_j(\mathbf{x})dt$ gives the transition probability of the j-th reaction to occur in the time interval $[t, t + dt]$. Propensities are typically dependent on the state of the system and the reaction conditions (i.e., external parameters) of the network such as temperature, pressure, etc.

Mathematically, $\{\mathbf{X}(t)\}_{t\in\mathbb{R}_+}$ is a continuous-time, time-homogeneous, pure jump Markov process with countable state space $E \subset \mathbb{N}^N$. The transition rates of the Markov process which completely define the process are given by the propensity functions $a_j(\cdot),\ j = 1, ..., M$. Indeed, the transition rates determine the clock of the updates (or jumps) from a current state $\mathbf{x}$ to a new (random) state $\mathbf{x}'$ through the total rate $a_0(\mathbf{x}) := \sum_{j=1}^{M} a_j(\mathbf{x})$ which is the intensity of the exponential waiting (or sojourn) time for a jump from state $\mathbf{x}$. The transition probabilities for the embedded Markov chain $\left\{J_n\right\}_{n\geq 0}$ with $J_n := \mathbf{X}(t_n)$ provided the current state $\mathbf{x}$ are $p_j(\mathbf{x}) = \frac{a_j(\mathbf{x})}{a_0(\mathbf{x})}$. There are exact algorithms for the simulation of the reaction network such as stochastic simulation algorithm (SSA) of Gillespie [20, 32] or next reaction algorithm of Gibson and Bruck [23] as well as approximation algorithms such as τ-leap [24] and several variations of it [26, 27]. As a demonstration, given that the system is at the state $\mathbf{X}(t) = \mathbf{x}$ at time t, SSA computes the waiting time δt as a random number drawn from an exponential distribution with the total rate $a_0(\mathbf{x})$ as parameter while the R_{j^*} reaction occurs where $j^* \in \{1, ..., M\}$ is chosen such that $\sum_{j=1}^{j^*-1} p_j(\mathbf{x}) < u < \sum_{j=j^*}^{M} p_j(\mathbf{x})$ where u is a random number uniformly chosen in the interval $[0, 1]$. The new state is given by $\mathbf{X}(t+\delta t) = \mathbf{x}' = \mathbf{x} + v_{j^*}$.

Relative Entropy

Assume that two probability distributions (or more generally probability measures) $\mathcal{P}$ and $\tilde{\mathcal{P}}$ have corresponding probability densities $p = p(x)$ and $\tilde{p} = \tilde{p}(x)$. Then, the Relative Entropy or Kullback-Leibler divergence of $\mathcal{P}$ with respect to $\tilde{\mathcal{P}}$ is defined as [33, 34]

$$\mathcal{R}\left(\mathcal{P} \,|\, \tilde{\mathcal{P}}\right) := \int p(x) \log\left(\frac{p(x)}{\tilde{p}(x)}\right) dx \ . \tag{1}$$

In a more general setting, relative entropy is defined as $\mathcal{R}\left(\mathcal{P}\,|\,\tilde{\mathcal{P}}\right) := \int \log\left(\frac{d\mathcal{P}}{d\tilde{\mathcal{P}}}\right) d\mathcal{P}$ where $\frac{d\mathcal{P}}{d\tilde{\mathcal{P}}}$ is a function known as Radon-Nikodym derivative while the integration is performed with respect to the probability $\mathcal{P}$, [35]. Relative entropy has been utilized in a diverse range of scientific fields from statistical mechanics to coding in telecommunications (information theory) and finance, and it possesses the following three fundamental properties, [34], [35]:

(i) it is always non-negative,

(ii) it is equal to zero if and only if $\mathcal{P} = \tilde{\mathcal{P}}$ $\mathcal{P}$-almost everywhere, and,

(iii) $0 \leq \mathcal{R}\left(\mathcal{P}\,|\,\tilde{\mathcal{P}}\right) < \infty$ if and only if $\mathcal{P}$ and $\tilde{\mathcal{P}}$ are absolutely continuous with respect to each other.

From an information theory perspective, relative entropy measures the loss of information when $\tilde{\mathcal{P}}$ is utilized instead of $\mathcal{P}$, [34]. In other words, relative entropy measures the inefficiency of assuming a perturbed distribution $\tilde{\mathcal{P}}$ instead of assuming the true distribution $\mathcal{P}$. Therefore, even though not a metric, relative entropy is a suitable quantity for the assessment of parametric sensitivity analysis since the higher the relative entropy due to some perturbation the larger the information loss in this direction.

Moreover, relative entropy can provide an upper bound for a large family of observable functions through the Pinsker (or Csiszar-Kullback-Pinsker) inequality. The Pinsker inequality states that the total variation norm between $\mathcal{P}$ and $\tilde{\mathcal{P}}$ is bounded from above by the square root of the twice of the relative entropy [34]. However, we are usually interesting on observable functions such as mean populations, variances, correlations, etc. where the Pinsker inequality states that for any bounded functions f holds:

$$|\mathbb{E}_{\mathcal{P}}[f] - \mathbb{E}_{\tilde{\mathcal{P}}}[f]| \leq ||f||_{\infty}\sqrt{2\mathcal{R}\left(\mathcal{P}\,|\,\tilde{\mathcal{P}}\right)}\,. \tag{2}$$

An obvious conclusion that is immediately drawn from the above inequality is that if the (pseudo-)distance between two distributions as measured by $\mathcal{R}\left(\mathcal{P}\,|\,\tilde{\mathcal{P}}\right)$ is controlled, then the error between any two observables is also controlled accordingly.

Pathwise Relative Entropy and Relative Entropy Rate

Proceeding to the pathwise formulation of the relative entropy, we assume that the propensities functions depend on a parameter vector $\theta \in \mathbb{R}^K$ (i.e., $a_j(\mathbf{x}) \equiv a_j^{\theta}(\mathbf{x})$) while the Markov process $\left\{\mathbf{X}(t)\right\}_{t\in\mathbb{R}_+}$ lies in the *stationary regime*. We denote by $\mu^{\theta}(\mathbf{x})$ the steady state (or stationary) distribution of the stochastic process $\mathbf{X}(t)$. The stationary path distribution of the process in the interval $[0,T]$ is denoted by $Q^{\theta}_{[0,T]}$. Notice that path distributions (i.e., time-series distributions) are high-dimensional complex objects; for instance, if we consider the simpler discrete-time Markov chain case $\{\mathbf{X}_n\}_{n\in\mathbb{Z}_+}$, defined by the transition probability density $p(x,x')$, then, utilizing repeatedly the Markov property, the stationary path distribution of a time-series (path) $(x_0, x_1, \ldots, x_T)$ is given by

$$Q_{[0,T]}(\{\mathbf{X}_n = \mathbf{x}_n\}_{0\leq n\leq T}) = \text{Prob}(\mathbf{x_0}, \ldots, \mathbf{x_T}) = \mu(\mathbf{x}_0)p(\mathbf{x}_0,\mathbf{x}_1)\ldots p(\mathbf{x}_{T-1},\mathbf{x}_T)\,. \tag{3}$$

Furthermore, we consider the jump Markov process $\{\tilde{\mathbf{X}}(t)\}_{t\in\mathbb{R}_+}$ defined by perturbing the propensity functions by a small vector $\epsilon \in \mathbb{R}^K$. The corresponding steady state and path distributions of $\{\tilde{\mathbf{X}}(t)\}_{t\in\mathbb{R}_+}$ are denoted by $\mu^{\theta+\epsilon}(\mathbf{x})$ and $Q^{\theta+\epsilon}_{[0,T]}$, respectively. Let the two path distributions $Q^{\theta}_{[0,T]}$ and $Q^{\theta+\epsilon}_{[0,T]}$ be absolutely continuous with respect to each other which is satisfied when $a_j^{\theta}(\mathbf{x}) = 0$ if and only if $a_j^{\theta+\epsilon}(\mathbf{x}) = 0$ holds for all $\mathbf{x} \in E$ and $j = 1, ..., M$. Then, the Relative Entropy of path distribution $Q^{\theta}_{[0,T]}$ with respect to $Q^{\theta+\epsilon}_{[0,T]}$ is defined similarly to (1) as

$$\mathcal{R}\left(Q^{\theta}_{[0,T]}\,|\,Q^{\theta+\epsilon}_{[0,T]}\right) := \int \log\left(\frac{dQ^{\theta}_{[0,T]}}{dQ^{\theta+\epsilon}_{[0,T]}}\right) dQ^{\theta}_{[0,T]}\,, \tag{4}$$

where $\frac{dQ^{\theta}_{[0,T]}}{dQ^{\theta+\epsilon}_{[0,T]}}$ is the Radon-Nikodym derivative of $Q^{\theta}_{[0,T]}$ with respect to $Q^{\theta+\epsilon}_{[0,T]}$. In fact, by Girsanov's formula, we can obtain an explicit expression for the Radon-Nikodym derivative in terms of the propensity rates, [35].

In the context of sensitivity analysis, the higher the relative entropy towards a direction ϵ, the larger the information loss t this direction, thus, the path distribution is more sensitive at this direction. Additionally, time-series distributions contain all the information about the stochastic process including the steady state distribution, as well as any type of observable, thus, pathwise relative entropy can be viewed as a "conservative" bound for sensitivity analysis. Indeed, having also in mind the Pinsker inequality (2), if relative entropy is insensitive to a parameter direction then any observable f is also insensitive to this direction.

Moreover, in the stationary regime, relative entropy increases linearly with time, hence, Relative Entropy Rate (RER)

$$\mathcal{H}\left(Q^{\theta} \,|\, Q^{\theta+\epsilon}\right) := \lim_{T\to\infty} \frac{1}{T}\mathcal{R}\left(Q^{\theta}_{[0,T]} \,|\, Q^{\theta+\epsilon}_{[0,T]}\right) \tag{5}$$

which is the time average of the relative entropy is well-defined, [36]. Due to (5), $\mathcal{H}\left(Q^{\theta} \,|\, Q^{\theta+\epsilon}\right)$ is an appropriate (time-independent) quantity for the measurement of sensitivity: it measures the rate of the loss of information due to an ϵ-perturbation of the model parameters, in the long-time, stationary dynamics regime of the stochastic process, as first proposed in [19]. Furthermore, RER admits an explicit formula given by

$$\mathcal{H}\left(Q^{\theta} \,|\, Q^{\theta+\epsilon}\right) = \mathbb{E}_{\mu^{\theta}}\Big[\sum_{j=1}^{M} a_j^{\theta}(\mathbf{x}) \log \frac{a_j^{\theta}(\mathbf{x})}{a_j^{\theta+\epsilon}(\mathbf{x})} - (a_0^{\theta}(\mathbf{x}) - a_0^{\theta+\epsilon}(\mathbf{x}))\Big]\,. \tag{6}$$

Thus, from a practical point of view, RER is an observable of the stochastic process which can be computed numerically as an ergodic average in a straightforward way since it only requires the known propensity functions. Nevertheless, in order to carry out the sensitivity analysis at the parameter vector θ, the computation of RER for different ϵ's is necessary, which for a high-dimensional parameter space may be computationally challenging. Thus, a sensitivity analysis methodology which does not depend on ϵ's –such methods are called gradient-free– is desirable and is developed next.

Pathwise Fisher Information Matrix

Even though not directly evident from (6), RER is locally a quadratic function of the parameter vector $\theta \in \mathbb{R}^K$. Indeed, RER is non-negative when $\epsilon \neq 0$ and equal to zero when $\epsilon = 0$ thus the linear order term is zero. Therefore, RER is written under a smoothness assumption on the propensity functions with respect to the parameter vector θ as, [19]:

$$\mathcal{H}\left(Q^{\theta} \,|\, Q^{\theta+\epsilon}\right) = \frac{1}{2}\epsilon^T \mathbf{F}_{\mathcal{H}}(Q^{\theta})\epsilon + O(|\epsilon|^3) \tag{7}$$

where $\mathbf{F}_{\mathcal{H}}(Q^{\theta})$ is a $K \times K$ matrix that can be considered as a pathwise analogue for the classical Fisher Information Matrix (FIM). Indeed, as in the classical FIM for parametrized distributions [34], (7) is the Hessian of the RER which geometrically corresponds to the curvature around the minimum value of the relative entropy rate. The pathwise FIM contains up to second-order accuracy all the sensitivity information for the path distribution at θ for any perturbation direction ϵ. This practically means that the computation of FIM is sufficient up to second order for the evaluation of all the sensitivities of the path distribution around the parameter vector θ.

The FIM can be utilized not only for the estimation/approximation of RER via (7) but also to infer intrinsic knowledge of the system's sensitivities [16, 37]. The spectral analysis of the FIM at θ reveals the (local) most/least sensitive directions of the system. Indeed, by ordering the eigenvalues of the FIM as

$$\lambda_1^{\theta} \geq ... \geq \lambda_K^{\theta}\,,$$

it can be inferred that the most sensitive direction corresponds to the eigenvector with eigenvalue λ_1^{θ} while the least sensitive direction corresponds to the eigenvector with eigenvalue λ_K^{θ}. Additionally, the classical FIM is one of the most useful tools for optimal experimental design. Many of the optimality criteria such as D-optimality where the determinant of the FIM is maximized or A-optimality where the trace of the inverse of the FIM are based on FIM, [37]. Even though the proposed pathwise FIM is not derived from a statistical description of measured data, meaning that the measurement noise is not taken into account, but from

the stochastic model itself, it can be utilized for the optimal design of experiments to account the intrinsic properties of the system. In the same direction, robustness of the system to parameter perturbations or errors as well as parameter identifiability can be studied utilizing spectral analysis of the proposed FIM. For instance, parameter identifiability is satisfied when all the eigenvalues of the FIM are above a given threshold, [18].

Moreover, we have an explicit formula for the pathwise FIM which is given by

$$\mathbf{F}_{\mathcal{H}}(Q^{\theta}) := \mathbb{E}_{\mu^{\theta}}\left[\sum_{j=1}^{M} a_j^{\theta}(\mathbf{x})\nabla_{\theta}\log a_j^{\theta}(\mathbf{x})\nabla_{\theta}\log a_j^{\theta}(\mathbf{x})^T\right] . \tag{8}$$

The implications of this explicit formula are twofold. First, it reveals that for many typical reaction networks the FIM has a special block-diagonal structure which reflects the parameter dependencies of the parameters and it is discussed in detail below. Second, the FIM is based on the propensity functions as well as on their derivatives which are assumed known –actually, they define the process– thus the FIM, similarly to RER, is numerically computable as an observable of the process. Subsequent sections present various strategies to numerically estimate in an efficient fashion both the RER and the FIM.

Sensitivity analysis on the logarithmic scale

In many biochemical reaction networks, the model parameters differ by orders of magnitude and the only meaningful option for sensitivity analysis is to perform parameter perturbations relative to the overall magnitude. In our setup, this can be done by perturbing the logarithm of the model parameters instead of the parameters itself. Thus, utilizing the chain rule for $\nabla_{\log\theta} f(\theta) = \nabla_{\theta} f(\theta).\nabla_{\log\theta}\theta = \theta.\nabla_{\theta} f(\theta)$ where '.' means element by element multiplication (i.e., $(a.b)_k = a_k b_k,\ k = 1, ..., K$), the logarithmically-scaled Fisher information matrix has elements:

$$\left(\mathbf{F}_{\mathcal{H}}(Q^{\log\theta})\right)_{k,l} = \theta_k\theta_l\left(\mathbf{F}_{\mathcal{H}}(Q^{\theta})\right)_{k,l} , \qquad k, l = 1, ..., K , \tag{9}$$

where $\mathbf{F}_{\mathcal{H}}(Q^{\theta})$ is given by (8). Similarly, the logarithmic perturbation for the RER is performed by utilizing the perturbation vector $\theta.\epsilon$ instead of ϵ. Notice that (7) continues to be valid for the logarithmic scale. Indeed, it holds that

$$\mathcal{H}\left(Q^{\theta} \,|\, Q^{\theta.(1+\epsilon)}\right) = \frac{1}{2}\epsilon^T\mathbf{F}_{\mathcal{H}}(Q^{\log\theta})\epsilon + O(|\epsilon|^3) . \tag{10}$$

Block-diagonal structure of FIM

In chemical reaction networks, reactions typically depend only on a small subset of the parameter vector. Mathematically, this is described as

$$a_j^{\theta}(\mathbf{x}) = a_j(\mathbf{x}; \theta_{k_1}, ...\theta_{k_{L_j}}) , \tag{11}$$

where $k_1, ..., k_{L_j} \in \{1, ..., K\}$ while $L_j \ll K$ is the number of involved parameters in R_j. Using (8), it can be shown that the parametric dependences of the propensity functions are inherited to the FIM. Indeed, after grouping the reactions into subsets in such a way that each subset contains the minimum number of reactions having common parameters, the FIM –upon rearrangement of the parameter vector–, becomes a block-diagonal matrix. The FIM is then written as

$$\mathbf{F}_{\mathcal{H}}(Q^{\log\theta}) = \begin{bmatrix} A_1^{\theta} & & 0 \\ & \ddots & \\ 0 & & A_M^{\theta} \end{bmatrix} \tag{12}$$

where $A_1^{\theta}, ..., A_M^{\theta}$ are block matrices. The block matrices which are defined by the reaction subsets with the same parametric dependence are easily obtained by creating a graph whose nodes are the reactions and the parameters and the edges are their dependences. Then, the parameter nodes contained in a connected subgraph define a parameter subset which in turn corresponds to a block of the FIM. An illustration of this procedure is shown in Figure 1 where a reaction network with $M = 9$ reactions and $K = 7$ parameters is plotted. The parametric dependencies of the reactions are shown in the left side where 4 subgroups of

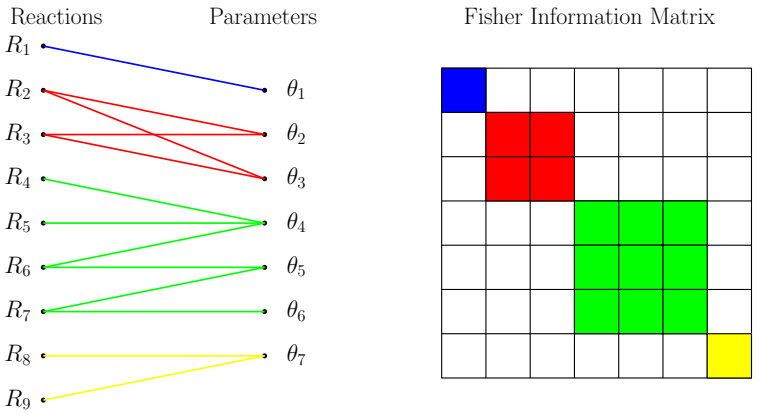

Figure 1: Left panel: The graph representation of the dependencies between the reactions (left column) and the model parameters (right column). The grouping of the parameters is then performed by pointing out the connected parts of the graph. Right panel: The corresponding block-diagonal structure of the FIM. In this example $K = 7$ while the largest dimension of the blocks is $L = 3$.

parameters are defined based on the graph connectivity. The resulting block-diagonal structure of the FIM is shown in the right hand side of Figure 1.

Before proceeding with the theoretical computation of the FIM for various well-known classes of biochemical reaction networks, we list some of the implications of this simplified structure of the FIM in sensitivity analysis and elsewhere.

(i) The sparsity of the FIM is proportional to the parametric decoupling between the reactions. Knowing a priori the zero elements of the FIM, there is no need to numerically compute them. It is clear that the computation cost for each sample drops from $O(K^2)$ to $O(KL)$ where L is the largest dimension of the block matrices.

(ii) The inverse of the FIM is also block-diagonal and each block of the inverse FIM is the inverse of the respective block of the forward FIM. This fact can be very useful for the estimation of a lower bound of the variance, i.e., Cramer-Rao bounds, [38,39] which is given by the diagonal elements of the inverse of the FIM.

(iii) From an optimal experiment point of view, [37], and based on these new insights on the FIM structure, the various optimality criteria are simplified. Indeed, the determinant utilized in the D-optimality test is the product of the determinants of the block matrices (mathematically, $\det(\mathbf{F}_{\mathcal{H}}) = \prod_{i=1}^{M} \det(A_i)$) while the trace of the inverse of the FIM utilized in the A-optimality test is the sum of the traces of the block inverses (i.e., $\mathrm{tr}({\mathbf{F}_{\mathcal{H}}}^{-1}) = \sum_{i=1}^{M} \mathrm{tr}(A_i^{-1})$).

(iv) It can provide insights on how to increase the identifiability of the parameters. For instance, the identifiability of the k-th parameter in the subsequent example will be increased if there is a way to increase the k-th diagonal element of the FIM. Analogously, the robustness of the system to the k-th parameter in the same example will be increased if the k-th diagonal element of the FIM is decreased accordingly.

Next we discuss two specific examples of biochemical reaction networks where the explicit calculation of the block-diagonal FIM is performed.

Reactions with independent reaction constants

An important class of well-mixed reaction networks has reactions being in the general form "$\alpha_j A_j + \beta_j B_j \xrightarrow{\theta_j} \ldots$" where A_j and B_j are the reactant species while α_j and β_j are the respective number of molecules needed for the occurrence of the reaction. The reaction constant, θ_j, is the parameter of the j-th

reaction. The propensity function for the j-th reaction is given as the product between a rate constant and a function of the current state $\mathbf{x}$:

$$a_j(\mathbf{x}) = \theta_j g_j(\mathbf{x}), \quad j = 1, ..., M \ . \tag{13}$$

Typically, $g_j(\mathbf{x}) = \mathbf{x}_A^{\alpha}\mathbf{x}_B^{\beta}/(\alpha!\beta!)$, however, it can take different forms depending on the modeling of the physical process. Overall, the reaction network has $K = M$ parameters, while each propensity depends only on one parameter, i.e., $L_j = 1$ in (11) for $j = 1, ..., M$. Proceeding, the (k, l)-th element of the FIM in the logarithmic scale is given by

$$\left(\mathbf{F}_{\mathcal{H}}(Q^{\log\theta})\right)_{k,l} = \theta_k\theta_l\mathbb{E}_{\mu^\theta}\left[\sum_{j=1}^{M} a_j^\theta(\ fx)\partial_{\theta_k}\log a_j^\theta(\mathbf{x})\partial_{\theta_l}\log a_j^\theta(\mathbf{x})^T\right] \ , \tag{14}$$

where μ^θ is the stationary distribution of the stochastic process. However, it is easily obtained for propensity functions of type (13) that $\partial_{\theta_k}\log a_j^\theta(\mathbf{x}) = \frac{1}{\theta_k}\delta_k(j)$ where $\delta(\cdot)$ is the Dirac function, therefore the FIM is a diagonal matrix with elements given by

$$\left(\mathbf{F}_{\mathcal{H}}(Q^{\log\theta})\right)_{k,l} = \begin{cases} \mathbb{E}_{\mu^\theta}\left[a_k^\theta(\mathbf{x})\right] \ , & l = k \\ 0 \ , & l \neq k \end{cases} \tag{15}$$

The diagonal elements of FIM are simply the stationary average of the respective propensity function. This rather unexpected result implies that the sensitivity of a reaction constant is proportional to the equilibrium average of the respective propensity function. However, notice that even though $\mathbb{E}_{\mu^\theta}\left[a_k^\theta(\mathbf{x})\right] = \theta_k\mathbb{E}_{\mu^\theta}\left[g_k^\theta(\mathbf{x})\right]$ the diagonal elements of the FIM are not linear functions of the corresponding reaction constants because the steady-state distribution μ^θ, depends also on the parameter vector θ.

Moreover, due to the diagonal form of the FIM, it is straightforward to perform eigenvalue analysis and make inference about the most/least sensitive directions of the reaction network. Actually, the eigenvalues of the FIM are its diagonal elements while the eigenvectors are the standard basis vectors of $\mathbb{R}^K$. Hence the most (resp. least) sensitive parameter is obtained from the largest (resp. smallest) diagonal element of the FIM (15). Furthermore, (15) demonstrates that the (local) robustness of the reaction network to a specific parameter is inversely proportional to the average propensity of the corresponding reaction. Finally, another observation stemming from (15) is that the larger the average of the propensity function the more frequent the respective reaction is fired, thus, it can be said that the faster reactions are more sensitive in a pathwise relative entropy sense.

Michaelis-Menten kinetics

Another important class of reaction networks is the Michaelis-Menten kinetics. This chemical network contains two reactions between species A and B (i.e., $A \leftrightarrow B$) with propensity functions given by

$$a_k^\theta(\mathbf{x}) = \frac{\theta_k\mathbf{x}_A}{\theta_{k+1} + \mathbf{x}_A} \qquad \text{and} \qquad a_{k+1}^\theta(\mathbf{x}) = \frac{\theta_k\mathbf{x}_B}{\theta_{k+1} + \mathbf{x}_B} \ .$$

This reaction sub-network which is derived under the quasi-steady-state assumption is one of the best-known models of enzyme kinetics in biochemistry. The parameter θ_k (usually denoted by $V_{\max}$) represents the maximum rate achieved by the system, at maximum (saturating) substrate concentrations while the Michaelis constant θ_{k+1} (usually denoted by K_m) is the substrate concentration at which the reaction rate is half the maximum value. The propensities of this Michaelis-Menten sub-network depend on two parameters ($L_k = L_{k+1} = 2$ in (11)) thus the corresponding FIM block is a 2×2 matrix. The elements of the FIM matrix again for the logarithmic scale are given by

$$\left(\mathbf{F}_{\mathcal{H}}(Q^{\log\theta})\right)_{k,l} = \begin{cases} \mathbb{E}_{\mu^\theta}\left[a_k^\theta(\mathbf{x}) + a_{k+1}^\theta(\mathbf{x})\right] \ , & l = k \\ -\mathbb{E}_{\mu^\theta}\left[a_k^\theta(\mathbf{x})\frac{\theta_{k+1}}{\theta_{k+1}+\mathbf{x}_A} + a_{k+1}^\theta(\mathbf{x})\frac{\theta_{k+1}}{\theta_{k+1}+\mathbf{x}_B}\right] \ , & l = k+1 \\ 0 \ , & l \neq k, k+1 \end{cases} \tag{16}$$

for the k-th row while the $k+1$-th row is given by

$$\left(\mathbf{F}_{\mathcal{H}}(Q^{\log\theta})\right)_{k+1,l} = \begin{cases} \mathbb{E}_{\mu^\theta}\left[a_k^\theta(\mathbf{x})\frac{\theta_{k+1}^2}{(\theta_{k+1}+\mathbf{x}_A)^2} + a_{k+1}^\theta(\mathbf{x})\frac{\theta_{k+1}^2}{(\theta_{k+1}+\mathbf{x}_B)^2}\right] \ , & l = k+1 \\ -\mathbb{E}_{\mu^\theta}\left[a_k^\theta(\mathbf{x})\frac{\theta_{k+1}}{\theta_{k+1}+\mathbf{x}_A} + a_{k+1}^\theta(\mathbf{x})\frac{\theta_{k+1}}{\theta_{k+1}+\mathbf{x}_B}\right] \ , & l = k \\ 0 \ , & l \neq k+1, k \ . \end{cases} \tag{17}$$

Strategies for the statistical estimation of RER and FIM

Previous sections introduced and justified RER and FIM as appropriate observables for measuring the sensitivity analysis of the reaction network's parameters. This section presents various strategies how to efficiently estimate these quantities as ergodic averages of the stochastic process.

Unbiased Statistical estimators

Since the stationary distribution, μ^θ, is usually not known, both FIM and RER should be estimated numerically as ergodic averages. Indeed, the statistical ergodic estimator for RER is given by

$$\bar{\mathcal{H}}^{(n)} = \frac{1}{T}\sum_{i=0}^{n-1}\delta t_i\Big[\sum_{j=1}^{M} a_j^\theta(\mathbf{x}_i)\log\frac{a_j^\theta(\mathbf{x}_i)}{a_j^{\theta+\epsilon}(\mathbf{x}_i)} - \big(a_0^\theta(\mathbf{x}_i) - a_0^{\theta+\epsilon}(\mathbf{x}_i)\big)\Big] \tag{18}$$

where δt_i is an exponential random variable with parameter the total rate, $a_0^\theta(\mathbf{x}_i)$, while $T = \sum_{i=1}^n \delta t_i$ is the total simulation time. The sequence $\{\mathbf{x}_i\}_{i=0}^n$ is the embedded Markov chain with transition probabilities from state $\mathbf{x}_i$ to state $\mathbf{x}_{i+1}$ is given by $p^\theta(\mathbf{x}_i) = \frac{a_j^\theta(\mathbf{x}_i)}{a_0^\theta(\mathbf{x}_i)}$. Notice that the weight δt_i at each step which is the waiting time at state $\mathbf{x}_i$ is necessary for the unbiased estimation of the observable [32]. Similarly, the unbiased estimator for the FIM is

$$\bar{\mathbf{F}}_{\mathcal{H}}^{(n)} = \frac{1}{T}\sum_{i=0}^{n-1}\delta t_i\sum_{j=1}^{M} a_j^\theta(\mathbf{x}_i)\nabla_\theta\log a_j^\theta(\mathbf{x}_i)\nabla_\theta\log a_j^\theta(\mathbf{x}_i)^T\ . \tag{19}$$

Notice that the computation of the local propensity functions $a_j^\theta(\mathbf{x}_i)$ for all $j = 1, ..., M$ is also needed for the simulation of the jump Markov process when Monte Carlo methods such as SSAs [32] is utilized. Thus, the computation of the perturbed transition rates is the only additional computational cost for the numerical RER while the additional cost for the estimation of the FIM is the computation of the derivatives of the (log-)propensities. *Algorithm 1* summarizes the numerical computation of RER and FIM utilizing SSA for the simulation of the jump Markov process.

Algorithm 1: SSA-based numerical computation of RER and FIM.

1. Initialize: $\mathbf{x} = \mathbf{x}_0$, $T = 0$, $\bar{\mathcal{H}} = 0$ and $\bar{\mathbf{F}} = 0$.
2. FOR $i = 1, ..., n$
   (a) Compute: $\big\{a_j^\theta(\mathbf{x})\big\}_{j=1}^M$, $a_0^\theta(\mathbf{x})$. Compute also $\big\{a_j^{\theta+\epsilon}(\mathbf{x})\big\}_{j=1}^M$ (only for RER) and $\big\{\nabla_\theta\log a_j^\theta(\mathbf{x})\big\}_{j=1}^M$ (only for FIM).
   (b) $\delta t = log(u_1)/a_0^\theta(\mathbf{x})$ where $u_1 \sim \mathcal{U}(0,1)$.
   (c) Update time: $T = T + \delta t$
   (d) Update RER: $\bar{\mathcal{H}} = \bar{\mathcal{H}} + \delta t\Big[\sum_{j=1}^M a_j^\theta(\mathbf{x})\log\frac{a_j^\theta(\mathbf{x})}{a_j^{\theta+\epsilon}(\mathbf{x})} - \big(a_0^\theta(\mathbf{x}) - a_0^{\theta+\epsilon}(\mathbf{x})\big)\Big]$.
   (e) Update FIM: $\bar{\mathbf{F}} = \bar{\mathbf{F}} + \delta t\sum_{j=1}^M a_j^\theta(\mathbf{x})\nabla_\theta\log a_j^\theta(\mathbf{x})\nabla_\theta\log a_j^\theta(\mathbf{x})^T$.
   (f) Find j^* such that $\sum_{j=1}^{j^*-1} a_j(\mathbf{x}) < u_2 a_0(\mathbf{x}) < \sum_{j=j^*}^M a_j(\mathbf{x})$ where $u_2 \sim \mathcal{U}(0,1)$.
   (g) Update state: $\mathbf{x} = \mathbf{x} + \nu_{j^*}$.
3. Normalize: $\bar{\mathcal{H}} = \bar{\mathcal{H}}/T$ and $\bar{\mathbf{F}} = \bar{\mathbf{F}}/T$.

Accelerated statistical estimators

A typical feature of biochemical systems is that the modeled reaction network is large with hundreds or thousands of reactions with different time scales stemming from the orders of magnitude difference between the reaction rates and/or between the species concentrations, making the SSA extremely slow. A large number of multi-scale approximations of the original SSA have been developed in order to handle such issues resulting to accelerated simulation algorithms. For example, mean-field approximation ignores the fluctuations of the stochastic process and yields a deterministic ODE system for the mean population of the species [40, 41]. Stochastic corrections to the mean-field model such as stochastic Langevin [25] and linear noise approximation [42] can be applied in order to improve the accuracy of the simulation. An alternative approximation of the jump Markov process is the tau-leap method proposed by Gillespie [24] where a batch of events occurs at each time-increment, τ. Several improvements on the basic tau-leap algorithm on how to select adaptively the τ [43] or on avoiding negative populations [27, 44] have been proposed, however, their performance is heavily model dependent.

In this subsection, we suggest applying such accelerated approximations in order to efficiently compute the FIM and/or RER observables, while maintaining controlled bias in the statistical estimators. As an illustration, we present the well-established and well-known to biologists due to its computational efficiency mean-field approximation. To proceed, the stochastic process is written as

$$\mathbf{X}(t) = x(t) + \eta\xi(t) \tag{20}$$

where $x(t)$ is the deterministic part (mean) of the process, $\xi(t)$ is the stochastic zero-mean part while η is the amplitude of the stochastic term. The amplitude of the stochastic term is proportional to the inverse square root of the reactant populations [25,42,45]. Thus, for large populations, the fluctuations of the time-evolving species populations become vanishingly small compared to the deterministic contributions. Consequently, the dominant part of the process is the deterministic term whose dynamics are given by the ODE system

$$\dot{x}_i(t) = \sum_{j=1}^{M} v_{j,i} a_j^{\theta}(x(t)) \ , \quad i = 1, ..., N \ . \tag{21}$$

The ODE system (21) is also known as reaction rate equations [25]. Restricted for simplicity to the special case presented in Section 1, the diagonal elements of the FIM are approximated as

$$\begin{aligned} \left(\mathbf{F}_{\mathcal{H}}(Q^{\log\theta})\right)_{k,k} = \mathbb{E}_{\mu^{\theta}}\left[a_k^{\theta}(\mathbf{x})\right] &\approx \frac{1}{T}\sum_{i=1}^{n} \delta t_i a_k^{\theta}(\mathbf{X}(t_i)) \\ &= \frac{1}{T}\sum_{i=1}^{n} \delta t_i \, a_k^{\theta}\big(x(t_i) + \eta\xi(t_i)\big) \\ &= \frac{1}{T}\sum_{i=1}^{n} \delta t_i \, a_k^{\theta}(x(t_i)) + O(\eta) \end{aligned} \tag{22}$$

where we made use of (20). Moreover, the ODE system is assumed to be solved with an adaptive time-step numerical integrator up to final time $T = \sum_{i=0}^{n} \delta t_i$. Thus, for large species populations ($|S_i| \gg 1$), the following numerical estimator for the FIM's diagonal elements

$$\left(\bar{\bar{\mathbf{F}}}_{\mathcal{H}}^{(n)}\right)_{k,k} = \frac{1}{T}\sum_{i=1}^{n} \delta t_i \, a_k^{\theta}(x(t_i)) \quad , \ k = 1, ..., K \tag{23}$$

is derived. Relation (23) suggests an algorithm similar to *Algorithm 1* for the numerical computation of the FIM adapted, of course, to the deterministic case where instead of using SSA, an ODE solver is utilized.

Remark 1: All multi-scale approximations try to overcome the problem of slow evolution of the system due to large number of species as well as the stiffness of the reactions. They are valid for large populations and relatively simple systems which do not exhibit complex dynamics such as bistability or intermittency. Indeed, as large deviation arguments [46] or even explicitly available formulas for escape times [47] show that stochastic approximations cannot always capture correctly exit times, rare events, strong intermittency, etc. However, in order to simulate large biochemical systems there is no other alternative than such approximate models, which nevertheless need to be employed with the necessary caution.

Remark 2: In biochemical systems, scientists are interested not only on the steady state (i.e., stationary) regime (equilibrium or not) but also on the transient regime. For instance, signaling phenomena which crucially affect and determine the long time states of the system belong to the transient regime. The extension of the proposed sensitivity analysis method to the transient regime is justified by the fact that the time-normalized relative entropy can be also decomposed as a sum of simple integrals which results to the fact that the statistical estimators (18) and (19) remain correct and valid. Subsequent section extensively presents an example of a biochemical system (EGFR) which exhibits transient behavior, and where the proposed sensitivity analysis tools are tested and validated. The rigorous mathematical derivation of the relative entropy rate for the transient regime is out of the scope of this publication and a dedicated mathematical article on the time dependent relative entropy rate will follow.

Results

The p53 Gene Model

The p53 gene plays a crucial role for effective tumor suppression in human beings as its universal inactivation in cancer cells suggests [28, 48, 49]. The p53 gene is activated in response to DNA damage and it constitutes a negative feedback loop with the oncogene protein Mdm2. Models of negative feedback are capable of oscillatory behavior with a phase shift between the gene concentrations. Here, we perform sensitivity analysis to a simplified reaction network between three species p53, Mdm2-precursor and Mdm2 introduced in [28]. It consists of five reactions and seven parameters provided in Table 1. The nonlinear feedback regulator of p53 through Mdm2 takes place in the second reaction while the remaining four reactions fall in the special class where each reaction depends on one parameter. Due to these mechanisms a nontrivial steady state regime exists and can be characterized by stochastic oscillatory behavior, see for instance Figure 2. Hence the proposed RER and FIM methodology is directly applicable, and the corresponding FIM consists of 5 diagonal blocks with respective size 1×1, 3×3, 1×1, 1×1, 1×1. Furthermore, the sensitivity analysis of this model has been performed earlier in [18] based on a linear noise approximation. Here we present a detailed comparison between the two sensitivity analysis methodologies, since the one proposed here does not involve any approximation of the stochastic network dynamics.

Table 1: The reaction table with x corresponding to p53, y_0 to Mdm2-precursor while y corresponds to Mdm2. The state of the reaction model is defined as $\mathbf{x}=[x,y_0,y]^T$ while the parameter vector is defined as $\theta=[b_x,a_x,a_k,k,b_y,a_0,a_y]^T$.

Event	Reaction	Rate	Rate's derivative
R_1	$\emptyset\rightarrow x$	$a_1(\mathbf{x})=b_x$	$\nabla_\theta a_1(\mathbf{x})=[1,0,0,0,0,0,0]^T$
R_2	$x\rightarrow\emptyset$	$a_2(\mathbf{x})=a_x x+\frac{a_k y}{x+k}x$	$\nabla_\theta a_2(\mathbf{x})=[0,x,xy/(x+k),-a_k xy/(x+k)^2,0,0,0]^T$
R_3	$x\rightarrow x+y_0$	$a_3(\mathbf{x})=b_y x$	$\nabla_\theta a_3(\mathbf{x})=[0,0,0,0,x,0,0]^T$
R_4	$y_0\rightarrow y$	$a_4(\mathbf{x})=a_0 y_0$	$\nabla_\theta a_4(\mathbf{x})=[0,0,0,0,0,y_0,0]^T$
R_5	$y\rightarrow\emptyset$	$a_5(\mathbf{x})=a_y y$	$\nabla_\theta a_5(\mathbf{x})=[0,0,0,0,0,0,y]^T$

Figure 2 shows the time-series of the species for the parameter values in Table 2. Evidently, oscillatory behavior is observed at this parameter regime, where persistent random oscillations occur, ranging between high and low populations. On the other hand, the frequency of the oscillations is less variable as it has been already reported both experimentally and numerically [28]. Another interesting observation is that the concentration of p53 species usually hits the lower bound of its admissible value (populations cannot be negative) which results in stochastic effects far away from Gaussianity, as can be readily seen also in Figure 2.

In this case, we denote by $\theta=[b_x,a_x,a_k,k,b_y,a_0,a_y]^T$ the parameter vector. The numerical estimator for RER as well as the FIM in the logarithmic scale are computed utilizing *Algorithm 1.* Logarithmic sensitivity analysis is preferred because the range of the parameters values varies by orders of magnitude as can be seen in Table 2. The upper plot in Figure 3 shows the RER as a function of time for various perturbations. Viewing RER as an observable, it is striking the speed of relaxation of the estimator. Within two or three oscillation periods RER has been converged to its value even though the three species have significant oscillations and stochasticity, as Figure 2 shows. A major reason for the fast relaxation is the numerical estimator of RER where the summation is over all reactions even though only one reaction takes places at each jump (see (19)). Having the important property of quick convergence, global sensitivity analysis, where

not only a point of the parameter regime but also large subsets of the parameter space, can be efficiently performed [15]. The lower panel of Figure 3 shows the RER when only one of the parameters are perturbed by +10% or by -10%. Additionally, the RER computed from the FIM, utilizing (7), is also provided. The FIM approximation of RER is a second order approximation in terms of $|\epsilon|$, hence the computation of FIM is typically enough to fully resolve the local sensitivities of a model. Evidently, the most sensitive parameters here are b_x and a_k while the least sensitive parameters are a_x and k.

Table 2: Parameter values for the p53 model.

Parameter	b_x	a_x	a_k	k	b_y	a_0	a_y
Value	90	0.002	1.7	0.01	1.1	0.8	0.8

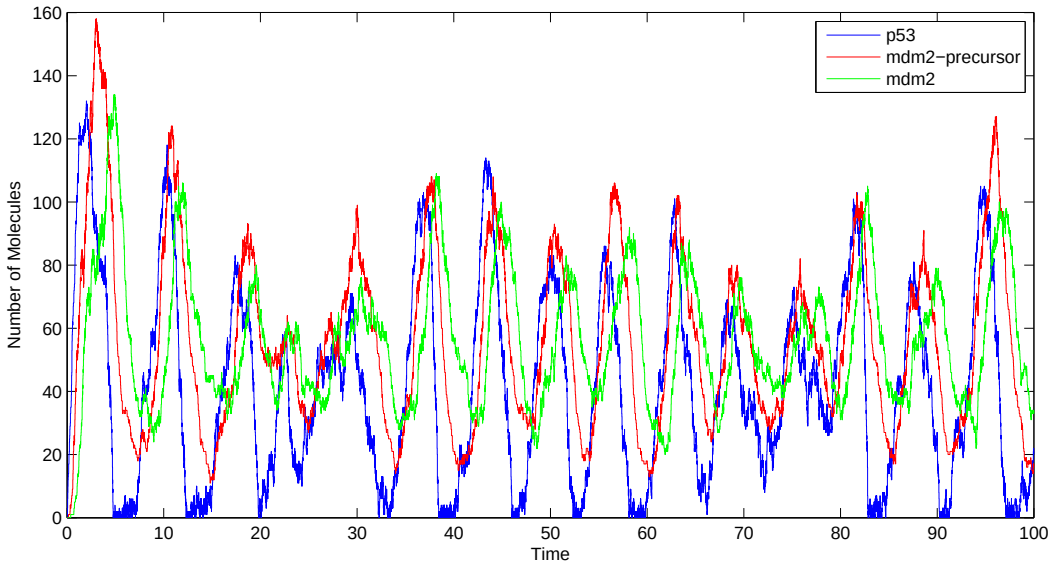

Figure 2: Molecule concentration of p53, Mdm2-precursor and Mdm2. Concentration oscillations as well as time delays (phase shifts) between the species are present due to the negative feedback loop. Furthermore, the concentration of p53 periodically approaches zero and since negative concentrations are not allowed, the stochastic characteristics of p53 are far from Gaussian.

Comparison with the LNA-based sensitivity approach

In [18], the authors suggested a linear noise approximation for the stochastic evolution around the non-linear mean field equation, and based on this approximation a system of ODEs is derived for the mean and the covariance matrix of the approximation process. Since the noise of LNA is Gaussian, the mean and the covariance matrix contains all the sufficient information about the approximate model. Then, the FIM is derived and based on it, the sensitivities for each parameter are computed. Although there are regimes where this approximation is applicable (short times, high populations, systems with a single steady state, etc.), for systems with nontrivial long-time dynamics, e.g. metastable, it is not correct as large deviation arguments [46] show, or even explicitly available formulas for escape times [47]. Similar issues with non-gaussianity in the long time dynamics arise in stochastic systems with strongly intermittent (pulse-like) or random oscillatory behavior [50]. In the p53 model considered in [18] which has the same parameter regime as here, Figure 2 reveals that the time-series of the p53 populations persistently fluctuates between high and

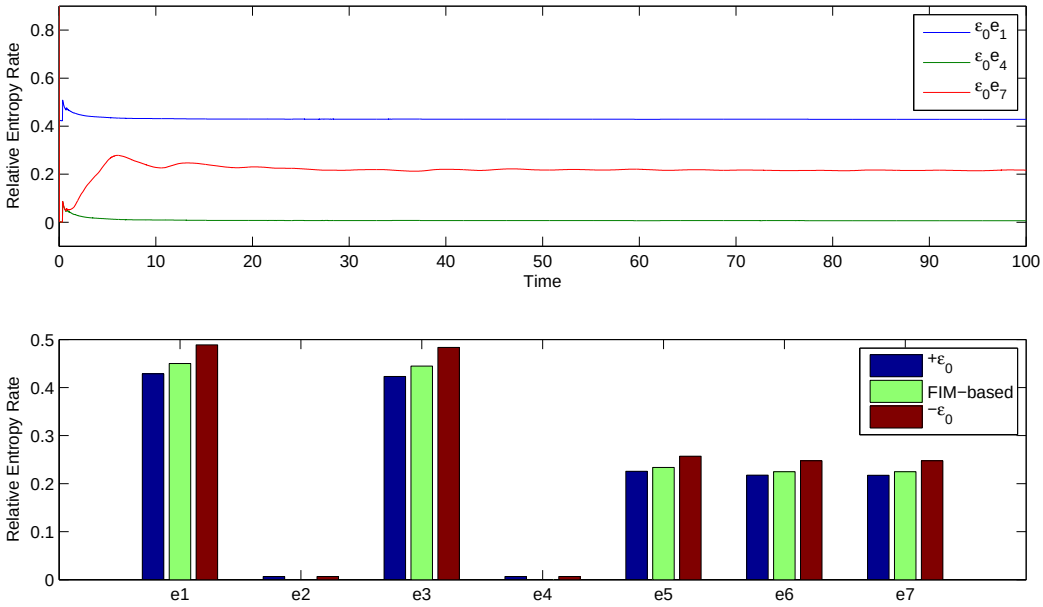

Figure 3: Upper panel: RER in time for the parameter perturbation of b_x (blue), k (green) and a_y (red) by $+10\%$ (i.e., $\epsilon_0 = 0.1$). The relaxation time of the RER as an observable is very fast. Lower panel: RER for various perturbation directions computed either directly (blue and red bars) or based on FIM (green bars).

low values, thus the LNA approximation may not be accurate at least when the concentration of the species is low.

At a coarse level, where the parameters are grouped into two classes depending on their sensitivities, the two sensitivity analysis approaches produce the same results. Indeed, by visual inspecting the lower plot of Figure 3 in the current publication and the Figure 3 in [18], the sensitive parameters in both methods are b_x, b_y, a_k, a_0, a_y while the insensitive parameters are a_x, k. However, upon closer inspection, the two methods produce very different results. Figure 4 shows the proposed FIM (left) based on the exact (without dynamics approximations) pathwise relative entropy theory, as well the FIM proposed in [18] which is derived from the LNA of the reaction system. The results are completely different and the proposed FIM based on path distributions is sparse as expected. A striking difference between the two sensitivity methods is that in our proposed method the sensitivity of parameter b_x is high while in the LNA-based method it is not (compare Figure 4 (dark blue) in this publication as well as Figure 3 in [18]). As a validation between the methods, we perturb b_x as well as b_y by the same amount and observe the time-series of the species. Figure 5 shows the time-series for the three species of the model for the unperturbed parameters (blue lines), the perturbation of only b_x by $+10\%$ (red lines) as well as the perturbation of only b_y by $+10\%$ (green lines). The same sequence of random numbers was used in all the simulations thus the time-series are visually comparable. It is evident that b_x is more sensitive than b_y as our sensitivity analysis method predicted while the LNA-based method suggested the reverse order of sensitivity. An explanation of the performance of the LNA-based sensitivity analysis stems from the fact that the p53 species does not have Gaussian noise when the population is close to zero and this can happen quite often as Figure 2 (blue line) shows. Additionally, notice that both b_x and b_y affect the concentration of p53 through the associated reactions thus their sensitivities are heavily biased due to the wrong statistical approximation of the p53 species.

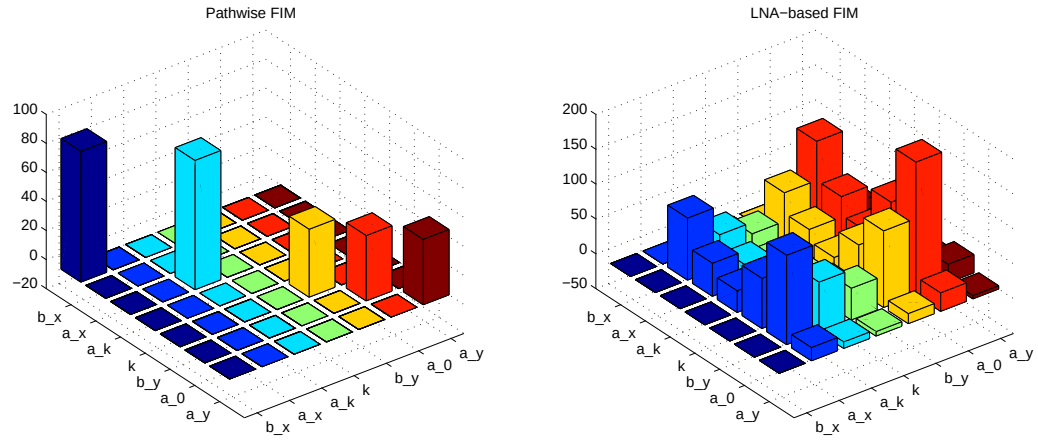

Figure 4: The proposed pathwise FIM (left) based on RER as well as the (scaled) FIM based on LNA computed from the StochSens package [51]. Evidently, the proposed method uncouples the parameter correlations since most of the off-diagonal elements are zero.

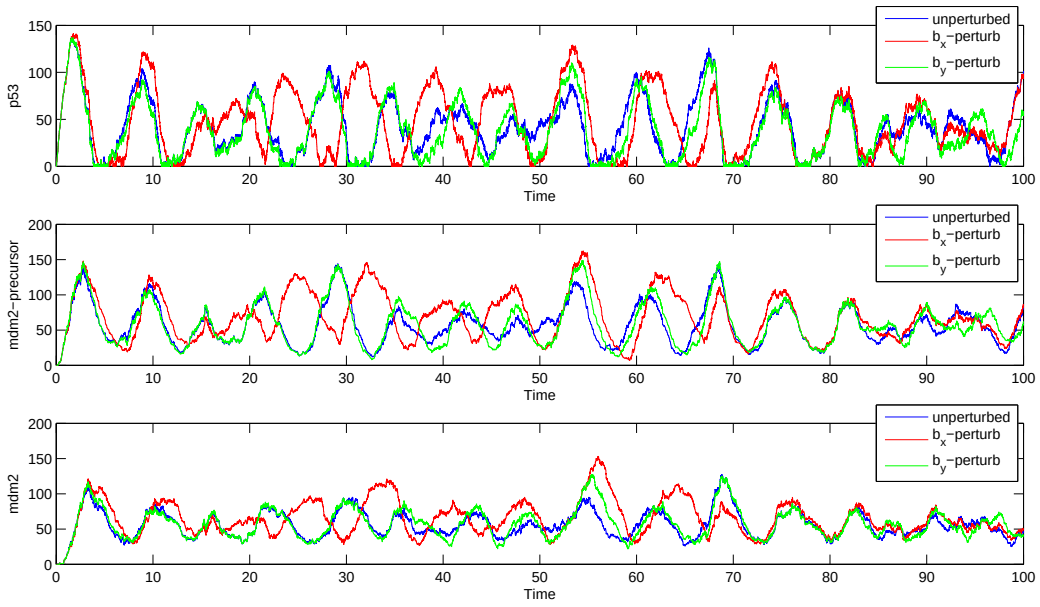

Figure 5: Time-series of the species in the p53 model for the unperturbed parameter regime (blue), when b_x is perturbed by +10% (red) as well as when b_y is perturbed by +10%. The same sequence of random numbers was used in all simulations. The visual comparison between the time-series suggests that the time-series properties are more sensitive to b_x than to b_y.

Epidermal Growth Factor Receptor model

The EGFR model is a well studied system describing signaling phenomena of (mammalian) cells [29–31]. As its name suggests, EGFR regulates cell growth, survival, proliferation and differentiation and plays a complex and crucial role during embryonic development and in tumor progression [52, 53]. In this paper, we adopt the reaction network for the dynamics of EGFR developed by Schoeberl et al. [31] which consists of 94 species with 207 reactions. Figure 6 presents at an abstract level the EGFR reaction network. Initially, the extracellular binding of EGF with the EGF receptors induce receptor dimerization. Then, two principal pathways, Shc-dependent and Sch-independend, are initiated leading to activation of Ras-GTP. Subsequently phosphorylation of MEK kinase through the activation of Raf kinase occurs leading to the phosphorylation of ERK kinase which regulates several proteins and nuclear transcription factors inside the cell. The detailed graphical description of the reaction network can be found in the Figures 1 & 2 of supplementary information in [31]. For the sake of completeness of this publication, all the reactions with their rates are provided in the supplementary files. The propensity functions for the reactions $R_1, ..., R_{97}, R_{100}, ..., R_{207}$ of the EGFR

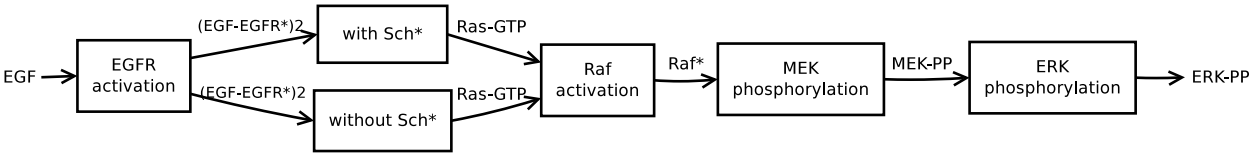

Figure 6: Building blocks of the EGFR reaction network. Each module communicates with the previous or next module through few species only. Additionally, except the first module, all the other modules are double, one external (i.e., outside the cell surface) and one internal.

network are written in the general form (see also (13))

$$a_j(\mathbf{x}) = k_j \mathbf{x}_{A_j}^{\alpha} \mathbf{x}_{B_j}^{\beta} / (\alpha! \beta!), \quad j = 1, ..., 97, 100, ..., 207 \tag{24}$$

with the exception of reaction pair R_{98}, R_{99} where their propensity functions are governed by the Michaelis-Menten kinetics

$$a_j(\mathbf{x}) = V_{\max} \mathbf{x}_{A_j} / (K_m + \mathbf{x}_{A_j}), \quad j = 98, 99 \tag{25}$$

where $\mathbf{x}$ is the current state of the reaction system while A_j corresponds to the reacting species. The parameter vector where sensitivity analysis will be performed is the reaction constants,

$$\theta = [k_1, ..., k_{97}, V_{\max}, K_m, k_{100}, ..., k_{207}]^T,$$

whose values are also provided in the supplementary txt file. Due to the specific values of the reaction constants as well as the initial population of the species (see Table 3), the firing rate between reactions differs by many orders of magnitude (highly stiff network), thus, even though there are few attempts to simulate the stochastic system [27], here for the purposes of RER and FIM calculations we adopt the mean-field deterministic approximation in the sense discussed in the accelerated estimators subsection. We solve the derived system of ODEs with Matlab's routine ode15s and compute the FIM at the steady state regime which corresponds to the time interval $[500, 700]$. As Figure 8 suggests, the completion of the internalization process needs about 500 seconds. It should be noted here that even though the simulation of the EGFR is performed utilizing a deterministic approximation model, the computed pathwise FIM has been derived from the *stochastic* network, (14). Therefore, under the validity of the approximation assumption (20), the computed FIM measures efficiently the sensitivities of the stochastic model in a gradient-free and sparse way.

Table 3: Initial population of the species for the EGFR network.

EGF	EGFR	GAP	Grb2	Sos	Ras-GDP	Shc
4.98e10	5e4	1.2e4	5.1e4	6.63e4	1.14e7	1.01e6
Raf	Phosphatase 1	Phosphatase 2	Phosphatase 3	MEK	ERK	Prot
4e4	4e4	4e4	1e6	2.2e7	2.1e7	8.1e4

Table 4: Ordering of the reaction rate constants based on their sensitivity index computed at the stationary regime. From left to right and from up to down.

k_{162}	k_{156}	k_{166}	k_{160}	k_{164}	k_{163}	k_{158}	k_{161}	k_{167}	k_{157}	k_{165}	k_{159}	k_{140}	k_{141}	k_{138}
k_{139}	k_{68}	k_{69}	k_{80}	k_{122}	k_{76}	k_{125}	k_{72}	k_{81}	k_{85}	k_{84}	k_{143}	k_{134}	k_{137}	k_{142}
k_{123}	k_{95}	k_{94}	k_{36}	k_{89}	k_{37}	k_5	k_6	k_{88}	k_{144}	k_{145}	k_{111}	k_{110}	k_{97}	k_{96}
k_{40}	k_{132}	k_{136}	k_{77}	k_3	k_4	k_{148}	k_{128}	k_{131}	k_{66}	k_{45}	k_{67}	k_{135}	k_{44}	k_{120}
k_{78}	k_{124}	k_{21}	k_{20}	k_{74}	k_{70}	k_{146}	k_{49}	k_{147}	K_m	k_{121}	k_{12}	k_{48}	k_{79}	k_{83}
k_{19}	k_{18}	k_{149}	k_{133}	k_{82}	k_{176}	k_1	k_{22}	k_{25}	k_{87}	k_{206}	k_{34}	k_{35}	k_{200}	k_{129}
k_{38}	k_{86}	k_{207}	k_{202}	k_{75}	k_{126}	k_{130}	k_{201}	k_{43}	k_{47}	k_{203}	k_{194}	k_{189}	k_{42}	k_{185}
k_{46}	k_{195}	k_{191}	k_{190}	k_{186}	k_{127}	k_{192}	k_2	k_{179}	k_{17}	k_{16}	k_{180}	k_{168}	k_{172}	k_{152}
k_{155}	k_{150}	k_{154}	k_{170}	k_{173}	k_{169}	k_{11}	k_{171}	k_{175}	k_{174}	k_{153}	k_{151}	k_{118}	k_{52}	k_{92}
k_{56}	k_{109}	k_{108}	k_{119}	k_{65}	k_{64}	$V_{\max}$	k_{73}	k_{53}	k_{32}	k_{33}	k_{103}	k_{102}	k_{41}	k_{116}
k_{50}	k_{90}	k_{54}	k_{117}	k_{205}	k_{63}	k_{62}	k_{198}	k_{51}	k_{93}	k_{199}	k_{101}	k_{30}	k_{31}	k_{100}
k_{193}	k_{187}	k_{184}	k_{71}	k_{188}	k_{91}	k_{181}	k_{178}	k_{39}	k_{61}	k_{57}	k_{10}	k_{114}	k_{106}	k_{60}
k_{24}	k_{13}	k_{29}	k_{59}	k_{55}	k_{28}	k_{112}	k_{104}	k_{23}	k_{204}	k_{196}	k_{107}	k_{58}	k_{14}	k_{115}
k_{27}	k_{197}	k_9	k_7	k_{26}	k_{183}	k_{182}	k_8	k_{105}	k_{113}	k_{15}	k_{177}	—	—	—

The upper plot of Figure 7 shows the diagonal elements of the FIM in descending order computed at the stationary (steady state) regime. The k-th diagonal element of the FIM corresponds to RER where the perturbation takes place only to the k-th parameter (see (7)). Figure 7 (upper plot) in conjunction with Table 4 where the reaction constants starting from the most sensitive towards the least sensitive parameter fully describe the (local) sensitivities of the reaction network. We report our results in the format of Figure 7 in order to be able to accommodate the large number of parameters in the model. Moreover, since the FIM is diagonal –except a small 2×2 block associated with the Michaelis-Menten reactions– the diagonal elements correspond to the eigenvalues of the FIM. The sensitivity analysis depicted in Figure 7, demonstrates that most model parameters allow for a vast range of perturbations without affecting the system's dynamics, see for instance Figure 8. Furthermore, this robustness to variations in most parameters was also reported in the original, fully deterministic EGFR model [31]. This is a feature shared by many multi-parameter models in systems biology and which is known as "sloppiness", [54]. Our methodology can easily demonstrate such properties in stochastic dynamics, as we can readily see in Figure 7, even if the models include a large number of parameters.

The earlier discussions refer to the analysis of the EGFR model close to steady state. On the other hand, EGFR is a signaling model whose transient regime, in addition to the steady state, is of great interest. As discussed in Remark 2 in Section 1, we can justify the application of the RER and FIM sensitivity analysis in the transient regime. Therefore, we compute the proposed FIM at the time interval $[0, 10]$, using (23). The lower plot of Figure 7 shows the diagonal elements of the FIM in the transient regime while keeping the ordering of the parameters unchanged as in the upper plot which depicts sensitivity at the steady state. The parameter sensitivities are completely different meaning that the sensitivities are time-dependent in the transient regime. For instance, the most sensitive parameters in the stationary regime correspond to the final products of the reaction network however in the time interval $[0, 10]$ these species have not been produced yet resulting to associated insensitive reaction constants.

Furthermore, the Pinsker inequality (2) implies that the insensitive parameters can be perturbed by orders of magnitude without affecting the species concentrations or any other observable. As an illustration of this fact, Figure 8 presents the concentrations of various critical species of the EGFR model when the 140-th (k_{65}) most sensitive parameter is perturbed, see Table 4. Solid blue lines correspond to the unperturbed parameter case while the dashed red lines correspond to the perturbed case where the perturbation is a multiple by a factor of ten of the corresponding reaction constant. We chose to present the total number of (EGF-EGFR*)2 binding species without Sch* (top, left panel) and with Sch* (top, middle panel) as well as Ras-GTP (top, right panel), total activated Raf or total Raf* (low, left panel), doubly phosphorated MEK or MEK-PP (low, middle panel) and doubly phosphorated ERK or ERK-PP. These species are important for the understanding of the system since the different modules of the EGFR reaction network communicate critically through these species (see Figure 6).

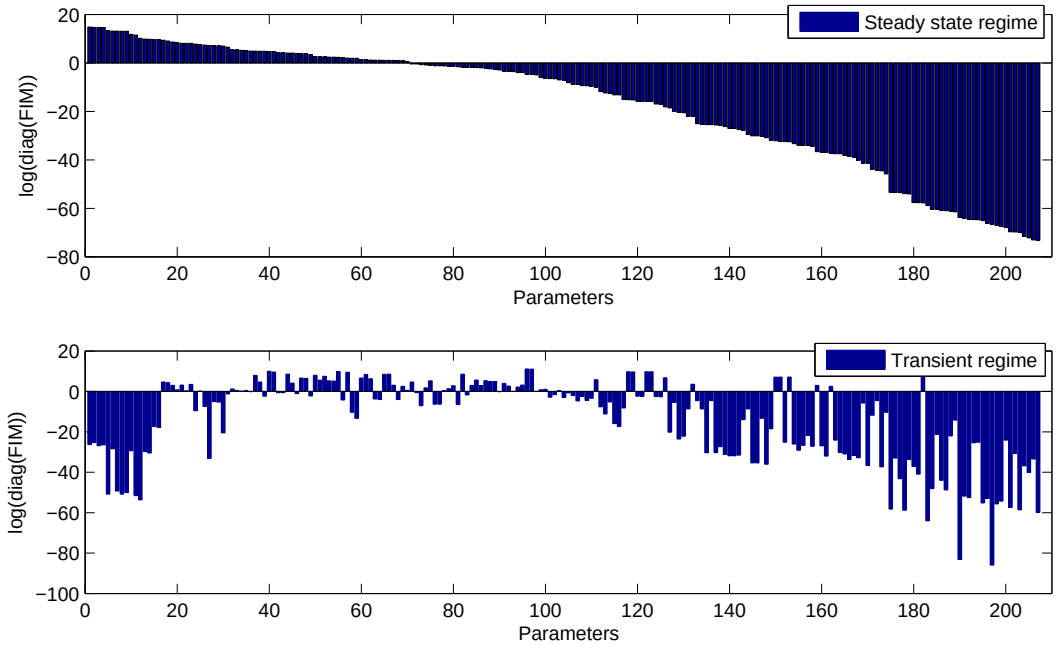

Figure 7: Diagonal elements of the FIM computed at the steady state regime (upper plot) and at the transient regime (lower plot). Presumably parameter sensitivities differ by orders of magnitude.

Conclusions

In this paper, we applied and extended a recently proposed parametric sensitivity analysis methodology to complex stochastic reaction networks. This sensitivity analysis approach is based on the quantification of information loss along different parameter perturbations between time-series distributions. This is achieved by employing the rigorously-derived Relative Entropy Rate, which is directly computable from the propensity functions only. A key aspect of the method is that we can derive rigorously a corresponding Fisher Information Matrix on path-space, which in turn constitutes a gradient-free approach to quantify the parametric sensitivity analysis; as such it provides a significant advantage in stochastic systems with a large number of parameters. We further demonstrated that the structure of the FIM revealed hidden, parameter dependencies and sensitivities between the reactions. The block-diagonal structure of the FIM highlighted the sparsity of the matrix which resulted in further increasing of the efficiency of the proposed method. Therefore, parametric sensitivity analysis for high-dimensional stochastic reaction systems is now tractable since it is well-known that in high dimensional stochastic systems the sensitivity analysis techniques can involve estimators of very high variance, e.g. in finite difference methods and their recently proposed variants, which can present an overwhelming computational cost [11]. Additionally, we proposed the use of multiscale numerical approximations of stochastic reaction networks in order to derive efficient statistical estimators for the FIM and tested one such approximation (mean field) to a high-dimensional system.

The proposed pathwise sensitivity analysis method is tested and validated on two biological systems: (a) the p53 reaction network where quasi-steady stochastic oscillations of the concentrations are observed and where multiscale approximations break down due to the consistent alternation between low and high populations, and (b) a stochastic epidermal growth factor receptor model which is an example of a high-dimensional reaction network with more than 200 reactions and a corresponding number of parameters. In the EGFR reaction network, we combined the proposed pathwise FIM which has been derived from the stochastic network and the mean field approximation which is used for the efficient estimation of the pathwise FIM. Moreover, our earlier rigorous analysis for the steady state regime [19] suggests suitable extensions in the transient regime which has been tested and validated for the EGFR model. We develop the full rigorous theory in an upcoming publication.

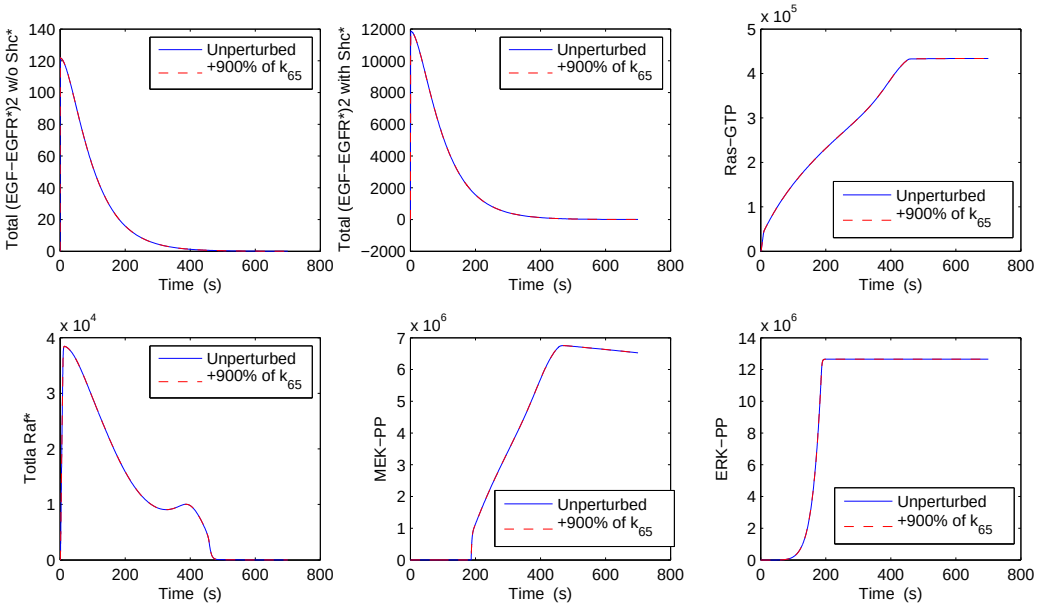

Figure 8: Time-dependent concentration of various species of the EGFR network either for the unperturbed parameter vector (solid blue lines) or for the perturbed one (dashed red lines). The 140-th most sensitive parameter (k_{65}) is ten-fold increased and the species concentrations are not affected. For the least sensitive parameters such as k_{65}, we rigorously know form Pinsker inequality (2) that they should not alter the concentration values or any other observable even when they are heavily perturbed.

Finally, we note that the relation between pathwise RER and various observables is not straightforward. However, we note that the path distribution contains all information regarding the process including the steady state and all time-dependent observables: practically, our proposed sensitivity analysis represents a conservative sensitivity estimate in the sense that insensitive directions for the relative entropy on path-space, will yield insensitive directions for every observable. This latter statement can be seen mathematically through the Pinsker inequality, (2). Based on these observations, the proposed sensitivity analysis methods can be easily used in complementary fashion with existing sensitivity analysis tools, as it can be used to narrow down the most sensitive directions in a system.

Acknowledgements

The work of MAK and YP was supported in part by the Office of Advanced Scientific Computing Research, U.S. Department of Energy under Contract No. de-sc0002339. The work of DGV was supported in part by the Office of Advanced Scientific Computing Research, U.S. Department of Energy under Contract No. DE-FG02-05ER25702. The work of MAK was also supported in part by the European Union (European Social Fund) and Greece (National Strategic Reference Framework), under the THALES Program, grant AMOSICSS.

References

1. Barkai N, Leibler S: **Robustness in simple biochemical networks**. *Nature* 1997, **387**(6636):913–917.
2. Csete M, Doyle J: **Reverse engineering of biological complexity**. *Science* 2002, **295**(5560):1664–1669.
3. Kitano H: **Opinion - Cancer as a robust system: implications for anticancer therapy**. *Nat. Rev. Cancer* 2004, **4**(3):227–235.

4. Donz A, Fanchon E, Gattepaille L, Maler O, Tracqui P: **Robustness Analysis and Behavior Discrimination in Enzymatic Reaction Networks**. *PLoS ONE* 2011, **6**:1–16.

5. Hart Y, Antebi Y, Mayo A, Friedman N, Alon U: **Design principles of cell circuits with paradoxical components**. *Proc. Nat. Acad. Sc. USA (PNAS)* 2012, **109**(21):8346–8351.

6. Glynn P: **Likelihood ratio gradient estimation for stochastic systems**. *Communications of the ACM* 1990, **33**(10):75–84.

7. Nakayama M, Goyal A, Glynn PW: **Likelihood Ratio Sensitivity Analysis for Markovian Models of Highly Dependable Systems**. *Stochastic Models* 1994, **10**:701–717.

8. Plyasunov S, Arkin AP: **Efficient stochastic sensitivity analysis of discrete event systems**. *J. Comp. Phys.* 2007, **221**:724–738.

9. Kim D, Debusschere B, Najm H: **Spectral Methods for Parametric Sensitivity in Stochastic Dynamical Systems**. *Biophysical Journal* 2007, **92**:379–393.

10. Rathinam M, Sheppard PW, Khammash M: **Efficient computation of parameter sensitivities of discrete stochastic chemical reaction networks**. *J. Chem. Phys.* 2010, **132**:034103–(1–13).

11. Anderson DF: **An efficient finite difference method for parameter sensitivities of continuous-time Markov chains**. *SIAM J. Numerical Analysis* 2012, **50**(5):2237–2258.

12. Sheppard P, Rathinam M, Khammash M: **A pathwise derivative approach to the computation of parameter sensitivities in discrete stochastic chemical systems**. *J. Chem. Phys.* 2012, **136**(3):034115.

13. McGill JA, Ogunnaike BA, Vlachos DG: **Efficient gradient estimation using finite differencing and likelihood ratios for kinetic Monte Carlo simulations**. *J. Comp. Phys.* 2012, **231**(21):7170–7186.

14. Liu H, Chen W, Sudjianto A: **Relative entropy based method for probabilistic sensitivity analysis in engineering design**. *J. Mechanical Design* 2006, **128**:326–336.

15. N Lüdtke and S Panzeri and M Brown and D S Broomhead and J Knowles and M A Montemurro and D B Kell: **Information-theoretic sensitivity analysis: a general method for credit assignment in complex networks**. *J. R. Soc. Interface* 2008, **5**:223–235.

16. Majda AJ, Gershgorin B: **Quantifying uncertainty in climate change science through empirical information theory**. *Proc. of the National Academy of Sciences* 2010, **107**(34):14958–14963.

17. Majda AJ, Gershgorin B: **Improving model fidelity and sensitivity for complex systems through empirical information theory**. *Proc. of the National Academy of Sciences* 2011, **108**(25):10044–10049.

18. Komorowski M, Costa MJ, Rand DA, Stumpf MPH: **Sensitivity, robustness, and identifiability in stochastic chemical kinetics models**. *Proc. Natl. Acad. Sci. USA* 2011, **108**:8645–8650.

19. Pantazis Y, Katsoulakis M: **A Relative Entropy Rate Method for Path Space Sensitivity Analysis of Stationary Complex Stochastic Dynamics**. *J. Chem. Phys.* 2013, **138**(5):054115.

20. Gillespie DT: **Exact Stochastic Simulation of Coupled Chemical Reactions**. *J. Chem. Phys.* 1977, **81**:2340–2361.

21. Chatterjee A, Vlachos DG: **An overview of spatial microscopic and accelerated kinetic Monte Carlo methods for materials' simulation**. *J. Computer-Aided Materials Design* 2007, **14**(2):253–308.

22. Slepoy A, Thompson A, Plimpton S: **A constant-time kinetic Monte Carlo algorithm for simulation of large biochemical reaction networks.** *J Chem. Phys.* 2008, **128**:(20):205101.

23. Gibson MA, Bruck J: **Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels**. *J. Chem. Phys.* 2000, **104**:1876–1889.

24. Gillespie DT: **Approximated accelerated stochastic simulation of chemically reacting systems**. *J. Chem. Phys.* 2001, **115**(4):1716–1733.

25. Gillespie DT: **The chemical Langevin equation**. *J. Chem. Phys.* 2000, **113**:297–306.

26. Rathinam M, Petzold LR, Cao Y, Gillespie DT: **Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method**. *J. Chem. Phys.* 2003, **119**:12784–12794.

27. Chatterjee A, Vlachos DG, Katsoulakis MA: **Binomial distribution based tau-leap accelerated stochastic simulation**. *J Chem. Phys.* 2005, **122**:024112.

28. Geva-Zatorsky N, Rosenfeld N, Itzkovitz S, Milo R, Sigal A, Dekel E, Yarnitzky T, Liron Y, Polak P, Lahav G, Alon U: **Oscillations and variability in the p53 system**. *Molecular Systems Biology* 2006, **2**:0033.

29. Moghal N, Sternberg P: **Multiple positive and negative regulators of signaling by the EGF receptor**. *Curr. Opin. Cell. Biol.* 1999, **11**:190–196.

30. Hackel P, Zwick E, Prenzel N, Ullrich A: **Epidermal growth factor receptors: critical mediators of multiple receptor pathways**. *Curr. Opin. Cell. Biol.* 1999, **11**:184–189.

31. Schoeberl B, C EJ, Gilles E, Muller G: **Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors**. *Nature Biotechnology* 2002, **20**:370–375.

32. Gillespie DT: **A general method for numerically simulating the stochastic time evolution of coupled chemical reactions**. *J. Comp. Phys.* 1976, **22**:403–434.

33. Kullback S: *Information theory and statistics*. John Wiley and Sons, NY 1959.

34. Cover TM, Thomas JA: *Elements of Information Theory*. Wiley Series in Telecommunications 1991.

35. Kipnis C, Landim C: *Scaling Limits of Interacting Particle Systems*. Springer-Verlag 1999.

36. Dumitrescu ME: **Some informational properties of Markov pure-jump processes**. *C. P. Matematiky* 1988, **113**:429–434.

37. Emery AF, Nenarokomov AV: **Optimal experiment design**. *Measurement Science & Technology* 1998, **9**:864–876.

38. Kay SM: *Funtamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall 1993.

39. Wasserman L: *All of Statistics: A Concise Course in Statistical Inference*. Springer 2004.

40. Gardiner C: *Handbook of Stochastic Methods: for Physics, Chemistry and the Natural Sciences*. Springer 1985.

41. van Kampen NG: *Stochastic Processes in Physics and Chemistry*. North Holland 2006.

42. Kurtz TG: **The Relationship between Stochastic and Deterministic Models for Chemical Reactions**. *J. Chem. Phys.* 1972, **57**:2976.

43. Cao Y, Gillespie DT, Petzold LR: **Efficient step size selection for the tau-leaping simulation method**. *J. Chem. Phys.* 2006, **124**:044109.

44. Tian T, Burrage K: **Binomial leap methods for simulating stochastic chemical kinetics**. *J. Chem. Phys.* 2004, **121**:10356.

45. Kurtz TG: *Approximation of population processes*. Society for Industrial and Applied Mathematics (SIAM) 1981.

46. Doering CR, Sargsyan KV, Sander LM, Vanden-Eijnden E: **Asymptotics of rare events in birth–death processes bypassing the exact solutions**. *Journal of Physics: Condensed Matter* 2007, **19**:065145–(1–12).

47. Hanggi P, Grabert H, Talkner P, Thomas H: **Bistable systems: Master equation versus Fokker-Planck modeling**. *Phys. Rev. A* 1984, **29**:371–378.

48. Prives C: **Signaling to p53: breaking the MDM2-p53 circuit**. *Cell* 1998, **95**:5–8.

49. Harris S, Levine A: **The p53 pathway: positive and negative feedback loops**. *Oncogene* 2005, **24**:899–908.

50. Katsoulakis MA, Majda AJ, Sopasakis A: **Intermittency, metastability and coarse graining for coupled deterministic-stochastic lattice systems**. *Nonlinearity* 2006, **19**(5):1021–1047.

51. Komorowski M, Zurauskiene J, Stumpf M: **StochSens–Matlab package for sensitivity analysis of stochastic chemical systems**. *Bioinformatics* 2012, **28**:731–3.

52. Sibilia M, Steinbach J, Stingl L, Aguzzi A, Wagner E: **A strain-independent postnatal neurodegeneration in mice lacking the EGF receptor**. *EMBO J.* 1998, **17**:719–731.

53. Kim H, Muller W: **The role of the EGF receptor family in tumorigenesis and metastasis**. *Exp. Cell Res.* 1999, **253**:78–87.

54. Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP: **Universally Sloppy Parameter Sensitivities in Systems Biology Models**. *PLOS Computational Biology* 2007, **3**.